



Les amorces de mots et leur contexte droit en français parlé spontané

Berthille Pallaud

► To cite this version:

Berthille Pallaud. Les amorces de mots et leur contexte droit en français parlé spontané. Travaux Interdisciplinaires sur la Parole et le Langage, 2005, 24, pp.117-138. hal-00136743

HAL Id: hal-00136743

<https://hal.science/hal-00136743>

Submitted on 16 Mar 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Les amorces de mots et leur contexte droit en français parlé spontané.

Berthille Pallaud

Parole et Langage, Université de Provence, Aix en Provence, France

Abstract

Parmi les phénomènes dits de disfluences qu'on peut identifier dans les énoncés de parole spontanée, amorces de mots et répétitions « involontaires » sont observées avec une grande fréquence. Traditionnellement, les amorces ont été analysées dans le cadre des répétitions disfluentes et donc confondues avec elles. Les observations faites sur ces deux types de disfluences nous ont conduits à les distinguer. Une étude spécifique des amorces de mots en français parlé a donc été entreprise depuis quelques années sur la base de corpus d'entretiens non directifs enregistrés et transcrits selon des conventions du Groupe Aixois de Recherche en Syntaxe (DELIC actuellement). L'objectif de notre étude est de décrire d'une part les insertions qui ont lieu à l'occasion des interruptions de mot et de leur réajustement et d'autre part, de préciser les types de reprises qui suivent ces interruptions ainsi que leur localisation. Certains paramètres comme la catégorie de troncation, le type de mots touchés par l'interruption et l'emplacement du fragment de mots par rapport au verbe recteur modulent la présence des insertions et les reprises après troncation. Il apparaît que les amorces de mots laissées inachevées sont, à quelques exceptions près, les seules, et une fois sur deux, suivies d'une rupture de construction verbale.

Mots-clefs :disfluences, amorces de mots, parole spontanée

Among the phenomena known as disfluencies which can be described in the statements of spontaneous spoken French, fragments of words and repetitions are observed with a great frequency. Traditionally, the word-fragments were analyzed within the framework of the disfluent repetitions and thus confused with them. The descriptions made of these two types of disfluences led us to propose distinguishing them. A specific study of the word-fragments in spoken French has thus been undertaken for a few years on non-directive talks corpora recorded and transcribed according to the French GARS team's conventions (DELIC currently). The objective of our study is to describe on the one hand, insertions which take place when the word is truncated and the type of re-adjustment, and on the other hand, to specify the types of repairs (and their localization) which follow these interruptions. Some parameters like the category of truncation, the type of words affected by the truncation and the site of the word-fragments (compared to the site of the verb) control the presence of insertions and the retracing after truncation. Moreover, the unfinished fragments of words appear to be the only one (with some rare exceptions), and once on two, to be followed by an interruption of the verbal construction.

Key-words: disfluencies, word-fragments, spoken French

1. Introduction

Si¹ la fluidité d'un énoncé oral se mesure à la régularité rythmique dans sa production, il est clair que les énoncés oraux ne sont pas fluides mais disfluent (Shriberg, 1999 ; Padeloup, 1992). Tout locuteur produit de la parole avec une certaine variabilité dans le débit des mots, des pauses

¹ Je remercie vivement Catherine Chanet, Danièle Duez et Marie-Laure Guénot pour leurs relectures attentives et pour leurs conseils.

(silencieuses ou non, Duez, 2001) et des allongements d'éléments linguistiques. Certains de ces éléments allongés sont observés au point d'interruption (IP) de l'énoncé que ce soit au milieu d'un mot (cas de l'amorce de mot) ou à la frontière de mots (ce qui est le cas du phénomène de bribe suivi ou non de répétition de mots ; Schriberg, 1999 ; Henry, 2002) :

ex. 1 Amorce de mot : ARBORIGN, 5, 14 c'est vrai que c'est pas **b-(IP) beau** d'associer les deux choses

ex. 2 Bribe suivie de répétition : Laurent, 1, 2 ils auront **leur propre(IP) leur propre** langage

Ces interruptions caractérisent, elles aussi, la disfluente dans la parole. Leur étude (en particulier, grâce au dispositif de la mise en grille ; Blanche-Benveniste, 1997), a moins insisté sur l'aspect disfluent de ces phénomènes que souligné la progression de l'énoncé par piétinements syntaxiques successifs (Jeanjean, 1984). Ainsi, ces « haltes » dans la production du texte sont obtenues par les pauses silencieuses ou remplies (nommées pauses d'hésitation, Duez, 1997) mais aussi par les répétitions et les amorces de mots.

Ces deux derniers types d'achoppements, cas particuliers des bribes, sont souvent traités de façon indistincte dans les études sur les disfluences (Candea, 2000, Clark & Wasow, 1998) ; les amorces de mots sont alors identifiées comme des répétitions partielles. Nous avons, au contraire, tenu à distinguer amorces et répétitions de mots en raison de leur grande différence de fréquence (deux fois plus de répétitions que d'amorces). Les résultats déjà obtenus confirment l'hypothèse ; ces phénomènes ne touchent ni les mêmes places syntaxiques ni les mêmes catégories grammaticales (Pallaud, 2003 ; Pallaud & Henry, 2004). Une analyse comparative des corpus (un locuteur par corpus) montre qu'il existe une grande hétérogénéité entre les locuteurs puisque les valeurs peuvent varier d'un facteur 10 (Pallaud, 2002). Cette variabilité observée dans la production des amorces suggère que ce paramètre reflète fortement une composante individuelle. Le phénomène d'amorce semble être une marque d'énonciation qui rend compte de caractéristiques subjectives en particulier d'hésitation. C'est si vrai que le terme d'hésitation en vient à définir de façon métonymique les disfluences : « Les hésitations peuvent fonctionner comme des marqueurs de personnalité » (Duez, 2001 p.129).

L'objectif de cette étude sur les amorces de mots est de préciser leur localisation syntaxique et le lieu de la reprise de l'énoncé quand il y en a une. Shriberg et Stolcke (1998) ont montré sur les répétitions (où, par définition, il y a toujours une reprise de l'énoncé) que le locuteur tend à remonter au début du constituant qu'il a du mal à reformuler. C'est une question qui sera envisagée à propos de l'amorce seule, du point de vue de sa position (le fragment et sa reprise) et de certains de ses aspects morpho-syntaxiques et syntaxiques.

2. Amorces et répétitions de mots : un rappel

2.1. Catégories d'amorces et types de mots

Trois catégories d'amorces ont été distinguées (Pallaud B., 2003) selon la place syntaxique occupée par ce qui suit l'amorce de mot. Lorsque l'élément suivant occupe la même place syntaxique, ce piétinement peut aboutir à compléter le fragment de mot (amorce complétée) ou ne pas le compléter mais le remplacer (amorce modifiée). Si l'élément qui suit l'amorce appartient à une autre place syntaxique, le fragment de mot ne sera ni complété ni remplacé (amorce inachevée). Un peu plus de la moitié des amorces de mots sont complétées (55% des cas) ; amorces modifiées (23% des cas) et inachevées (22% des cas) constituent le reste à part égale (Fig. 1)² :

ex.3 Amorce complétée : C5cBelfo, 10,2 je me suis toujours at- **attachée** à ce que les enfants

² Il est à noter que ces proportions obtenues sur un recueil de 441 fragments de mots sont les mêmes que celles que nous avons trouvées précédemment avec un recueil de 1000 fragments de mots (Pallaud, 2002a).

ex.4 Amorce modifiée : C6aBesan, 3 qui dit euh papier dit **évita- évidemment** crayon de papier qui va avec

ex.5 Amorce laissée inachevée : C7dBord, è, 15 ça ça a été pour moi le **pl-** et c'est toujours quelque chose

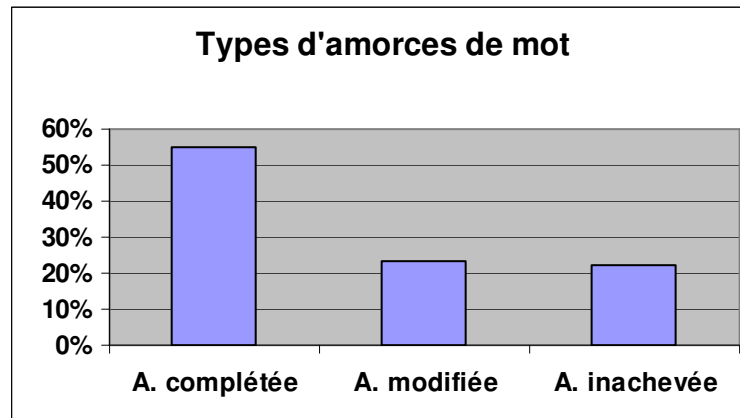


Figure 1 Répartition des amorces de mots

Les amorces complétées et les amorces modifiées sont les cas où le locuteur renoue ou « répare » le fil de son discours en reprenant une partie de son énoncé interrompu et en le poursuivant. Les fragments laissés inachevés sont ceux où il n'y a ni reprise ni réparation mais où le locuteur poursuit simplement son énoncé sans revenir en arrière.

Les proportions dans ces trois catégories suggèrent que ces troncations involontaires apparaissent dans au moins la moitié des cas comme des marques d'hésitation dans l'élaboration du texte (amorce complétée) que comme des manifestations d'une erreur à corriger (amorce inachevée et modifiée). De ce point de vue, notre observation confirme les études de Schegloff et al. (1977) sur les auto-corrections (qui correspondent dans notre étude à des amorces complétées,) et celles de Cappeau, (1998) sur les bribes (c'est à dire des amorces de syntagme).

Notre étude (Henry et Pallaud, 2003, 2004) a montré que les unités lexicales ne sont pas affectées de la même manière par les amorces et les répétitions. Il apparaît utile de distinguer les mots qui ont une charge lexicale pleine, les *mots pleins*, de ceux qui participent de la structuration de l'énoncé : les *mots-outils*. Les résultats obtenus pour les répétitions (Corpus *Corpaix* : 15 985 répétitions) et les amorces (999 amorces analysées) ont révélé une différence très nette (Fig. 2). Les répétitions touchent plus souvent (91,3%) les mots-outils que les amorces qui, elles, affectent plus fréquemment les mots pleins (67,9%). Les répétitions sont donc observées plus tôt que les amorces dans le syntagme puisque la catégorie des mots-outils est constituée principalement par les déterminants, les pronoms et les prépositions, items qui se trouvent en général en tête de syntagme.

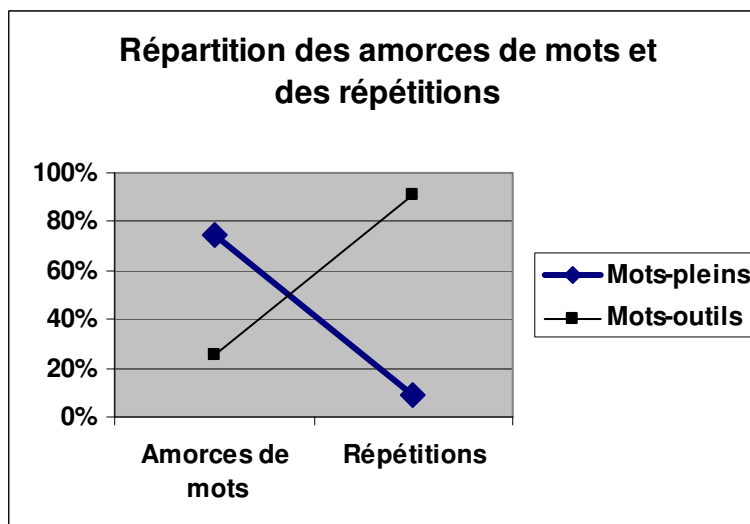


Figure 2 Répartition des amorces de mots et des répétitions.

Si on regarde (Fig.3) ce que deviennent ces deux types de mots lorsqu'ils sont tronqués, on constate qu'un mot-outil a autant de chance qu'un mot-plein d'être complété. Il n'en est pas de même pour les deux autres catégories de troncation : comparativement aux mots-pleins, les mots-outils sont plus souvent laissés inachevés et moins fréquemment modifiés. Cette caractéristique des fragments de mots-outils laissés inachevés sera examinée en fonction de ce qui suit ces amorces, en particulier, pour déterminer si ils sont suivis plus fréquemment d'une rupture syntaxique (conf plus loin).

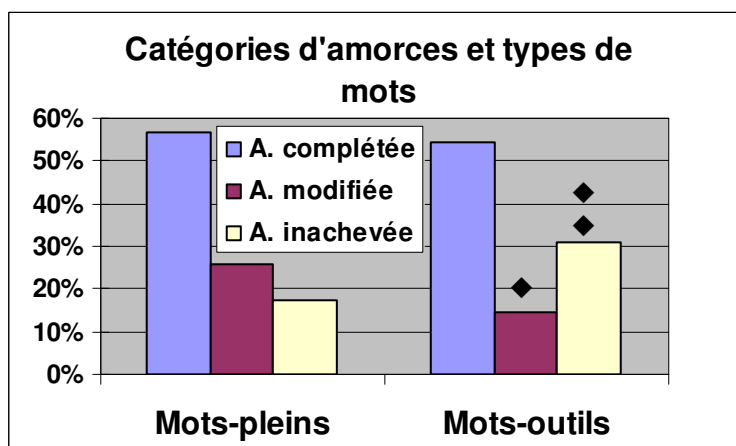


Figure 3 Types de mots et types de troncations

Nous avons également comparé la répartition des fragments de mot et des répétitions dans les zones de l'énoncé (avant le verbe, sur le verbe ou après le verbe). Là aussi la différence est nette entre ces deux types d'achoppement (Fig 4). Avant le verbe, l'hésitation est plus fréquemment marquée par une répétition que par une amorce de mot. C'est l'inverse après le verbe.

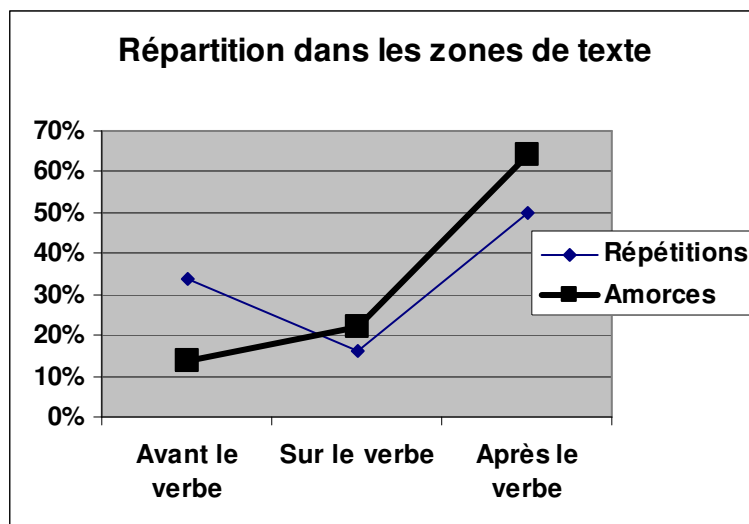


Figure 4 Répartitions des amorces de mots et des répétitions dans les zones de texte

Ainsi quel que soit le niveau considéré (celui du syntagme ou de la proposition), comparativement aux amorces, les répétitions se produisent plus tôt dans le syntagme et sont observées, également, plus tôt dans l'énoncé.

Cette approche contrastive (au regard des amorces et des répétitions) permet de faire des hypothèses notamment sur une fonction probable de la répétition (Henry et Pallaud, 2003, 2004) : à savoir permettre au locuteur par avancées successives de faire progresser son énoncé. L'amorce elle-même peut être répétée :

ex. 6 C7bBord, 10, 2 j'esp- j'esp- j'espère que j'ai pas été trop bavard

Lorsque l'amorce elle-même est répétée, elle est dans 82% des cas (au lieu de 59% pour les amorces simples, Fig. 5) complétée et peu modifiée ou inachevée.

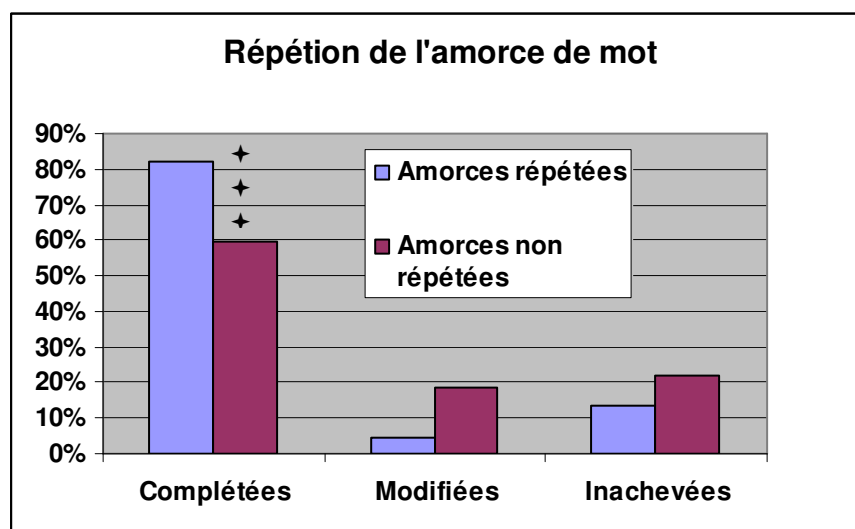


Figure 5 Comparaison des types d'amorcés lorsque le fragment est répété ou non.

Ce dernier résultat est très significativement différent³ de ce qui a été observé sur les amorces simples. La répétition de l'amorce paraît permettre plus souvent au locuteur de compléter le mot

³ ($\chi^2 = 29,85$; d.d.l.=2 ; $p < .001$)

qu'il vient de débiter. Ce constat a attiré l'attention sur le contexte ultérieur de l'amorce de mot et justifié une étude approfondie des reprises et réajustements après l'interruption de mot.

3. Corpus et méthodologie

L'étude porte sur un ensemble de 20 corpus recueillis (105 000 mots), à l'exception de l'un d'entre eux (CorpusEDF ; Henry et Pallaud 2004), dans le cadre d'une enquête menée en 1998 et 1999, par le Groupe Aixois de Recherche en Syntaxe (sous la responsabilité de Claire Blanche-Benveniste) sur le français parlé dans toutes les régions de France. La totalité de cette enquête a été poursuivie et numérisée (son et transcription) par l'équipe DELIC. Selon les conventions du Groupe Aixois de Recherche en Syntaxe qui prévoient une transcription orthographique des énoncés oraux, les amorces de mots sont notées par un trait d'union collé au fragment du mot et donc identifiable automatiquement de façon univoque.

Tous ces enregistrements sauf deux ont été conduits en privé, selon la méthode de l'interview la moins directive possible, et ne rassemblaient que deux locuteurs. Les deux corpus faits en public étaient des exposés improvisés devant un groupe de 40 personnes. Tous les locuteurs sont adultes.

Ces corpus ne sont pas étiquetés. L'analyse porte exclusivement sur les énoncés du locuteur interviewé. L'étiquetage (non automatique) est fait, dans un deuxième temps, sur les 441 données extraites de ces corpus (logiciel Contextes établi par J. Véronis) et rassemblées dans un tableur (Excel). Si on se fonde sur un débit moyen de 200 mots/min, le nombre total de mots étant de 105 000 mots, la durée totale de l'ensemble de ces corpus est de 7 heures 51 min. La longueur moyenne des corpus du GARS est de 3080 mots soit une durée moyenne de 16 min. ; les valeurs extrêmes de ces durées sont 1307 et 4931 mots. On retrouve, dans ce sous-ensemble de corpus, la fréquence moyenne d'apparition des amorces de mots dans un énoncé (1/57sec. ; variation individuelle de 1/23 sec. à 1/8 min.).

4. Les résultats: analyse des réajustements après les interruptions de mots

On constate que deux sortes de phénomènes de réajustement après les amorces de mots peuvent être décrits : d'une part, les **insertions** d'éléments linguistiques, d'autre part les **reprises** de l'énoncé (que ce soit par piétinement syntaxique ou par la poursuite de l'énoncé sur l'axe syntagmatique). Nous emploierons le terme de « reprise » uniquement dans le cas où le locuteur piétine sur la même place syntaxique c'est à dire reprend une partie de son énoncé avant de la poursuivre (ce qui correspond aux cas des amorces complétées ou modifiées).

4.1 Les insertions

La structure des achoppements (Levelt, 1983, Clark et Wasow, 1998 et Shriberg, 1999) qui distingue trois phases dans ces phénomènes de disfluente, fait apparaître deux espaces possibles d'insertion suite à la troncature: celui qui précède la reprise de l'énoncé et celui qui commence avec la reprise de l'énoncé.

Les trois phases de la structure sont les suivantes : **interruption, latence, reprise de l'énoncé** (schéma) :

*le **reparandum** (RM) : il désigne ici le fragment de mot. Le point d'interruption (IP) : établit la frontière finale du *reparandum* : (il est matérialisé par le trait d'union dans nos transcriptions).

*l'**interregnum** (IM, le *hyatus* chez Clark et Wasow, 1998) : désigne le moment compris entre la frontière finale du *reparandum* et la frontière initiale du *repar*.

*le **repair** ou **reparans** (RR) : représente la partie réparée, répétée ou modifiée du *reparandum* ; Deux lieux d'insertion (*en italique*) sont donc possibles à la suite de la troncature :

Exemple (C24Bnanc, 24,2) :

ça **dép-** (Reparandum RM, IP)

ouais [Espace IM]

ça **euh dépend** (Reparans, espace RR) + moi au début j'avais demandé à aller en au Népal

Deux sortes d'insertions ont été distinguées :

- des incises qui s'insèrent dans l'énoncé sans y être liées syntaxiquement (pauses voisées ou non, incises énonciatives, parenthétiques) (ex. 7 et 8).
- des incises liées syntaxiquement à l'énoncé interrompu et à sa reprise : les amorces répétées (espace IM uniquement ; ex. 9), des répétitions, des items ajoutés, enlevés ou remplacés (espace RR uniquement puisque la reprise est commencée ; conf. plus loin).

4.1.1 Le premier espace [espace IM "Interregnum"] n'est rempli que dans 13% des cas seulement.

On observe, toutes amorces de mots confondues, des insertions énonciatives (pauses voisées ou non, *enfin*, *ben*, *bon*, etc.), parenthétiques, et des répétitions du fragment de mot (ex.6) :

ex. 7 : EDF, 14,4 c'est qu'au premier choc euh **prétro-** [*euh*] **pétrolier** euh

ex.8 : C6bBesan 2,8, et j'ai **regard-** [*et j'ai en fait choisi ça par hasard*] j'ai **regardé** ce qu'il fallait avoir

ex.9 : C5aBelfo, 3,2, euh et **j'ai-** [*j'ai-*] **j'aimais** pas quoi

Le moment qui suit l'interruption de mot est donc bien un espace potentiel d'insertions énonciatives ou parenthétiques non liées syntaxiquement au contexte antérieur. Dans nos corpus, il n'est « utilisé » par le locuteur que dans un cas sur 10 ; dans les autres cas cette place reste vide. Cependant, la présence de ces insertions est dépendante du type d'amorces de mots. Près de la moitié (40%) de ces incises suivent un fragment de mot qui sera laissé inachevé.

La répartition de ces incises selon les types d'amorces (Fig 6) montre que la catégorie des amorces inachevées est spécifique à cet égard et diffère significativement⁴ des deux autres catégories. Si 26% des interruptions de mots laissés inachevés sont suivies d'une incise, cette proportion est deux fois moindre dans les catégories d'amorces où s'instaure un piétinement syntaxique. Ce résultat suggère que la présence d'une insertion dans l'espace IM devient indicatrice d'un événement ; elle se produit plus souvent après une interruption laissée inachevée et sera donc plus souvent suivie d'une poursuite de l'énoncé que d'un piétinement syntaxique. La production d'une insertion énonciative ou parenthétique prendrait alors la place d'un piétinement syntaxique ce qui permettrait au locuteur de poursuivre son énoncé sans avoir complété le syntagme.

⁴ Khi2= 14,45646 ; d.d.l. = 2 ; p<.001)

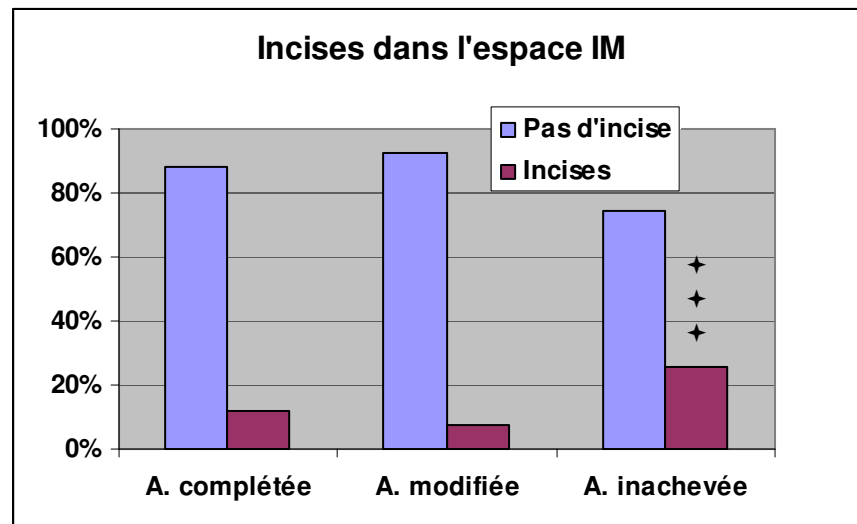


Figure 6 Présence des incises dans l'espace IM en fonction des types d'amorces

4.1.2. Le deuxième espace [espace RR, "Reparans"] débute avec la reprise de l'énoncé. Cette reprise restaure la continuité dans l'élaboration de l'énoncé qui avait été interrompu. On retrouve les mêmes insertions (non liées syntaxiquement) que celles décrites dans l'espace IM, mais elles sont beaucoup moins nombreuses cependant (3% seulement):

- (incise parenthétique) EDF, 18,5 pour euh mettre les **ni-** [les *comment* les **niveaux**] euh d'eau corrects

- (pause silencieuse) EDF, 13,2 mais la cogénération restera **u-** [une + une] **solution**

- (répétition) C24bNan, 2,1 c'était un peu euh un peu **com- comment comment** vivre au quotidien

Ce deuxième espace (RR), en revanche, est réservé à des insertions qui vont modifier l'énoncé (des « modifications antérieures ») interrompu par cette troncation. Grâce à ces « modifications antérieures », donc liées syntaxiquement au contexte antérieur, cet espace comporte, au total, deux fois plus d'insertions que l'espace IM (22% dans l'espace RR contre 13% dans l'espace IM).

Au total, un tiers des amorces de mots est donc suivi d'insertions diverses (dans les espaces IM et RR) et cela principalement dans la catégorie des amorces complétées. Comme le suggéraient nos observations sur les amorces répétées, ces insertions entourant surtout les amorces complétées semblent donc jouer un rôle dans la recherche lexicale en la facilitant. Ces deux espaces (Fig.7) ne comportent pas les mêmes types d'insertion. L'espace situé juste après la troncation semble le lieu des incises non liées syntaxiquement alors qu'elles sont presque absentes lorsque la reprise a commencée. Dans cet espace RR, la très grande majorité des incises modulent la reprise de l'énoncé et sont donc liées syntaxiquement au contexte antérieur.

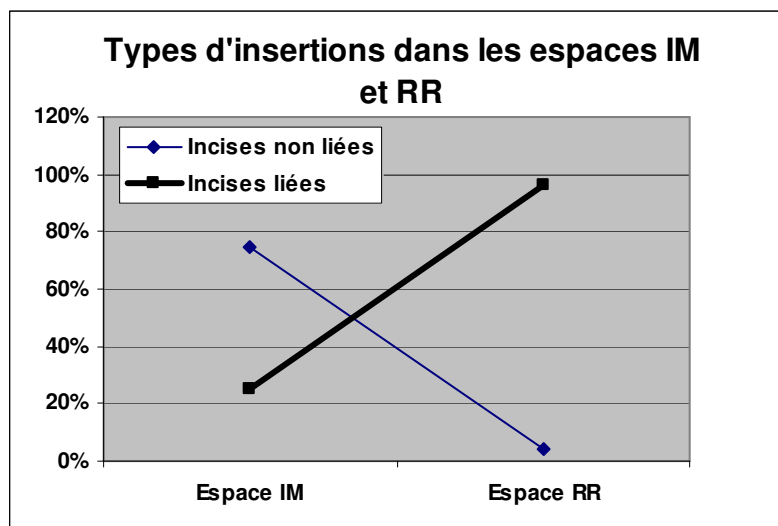


Figure 7 Types d'insertion liés ou non syntaxiquement à l'énoncé interrompu dans les espaces IM et RR.

Les insertions modulatrices de l'énoncé sont de trois sortes : des ajouts, des suppressions, des remplacements d'élément.

ex.10 : des incises qui modulent l'énoncé :

Item enlevé : C24bNan, 20, 1 il y **en** av- il y avait deux filles de dix-huit ans

Item ajouté : C7cBorde, 12, 11 le judoka commence à être pas mal et euh s'il a s'il s'en- **continue** à s'entraîner

Item remplacé : EDF 2,41 qui est un peu **plus** mo- un peu **moins** modulable

Ces catégories de modifications ne sont pas équivalentes puisque comme le montre le tableau suivant (1), elles varient dans leurs proportions. Le locuteur enrichit plutôt son énoncé ou remplace des éléments et dans une moindre mesure le réduit.

	Nombre	Pourcentage
item ajouté	35	54%
item enlevé	9	14%
Item remplacé	21	32%
Total	65	100%

Tableau 1 Types de modifications dans l'énoncé antérieur

Le fragment de mot n'est donc pas toujours le seul élément à observer. Des modifications peuvent avoir lieu ailleurs que sur ce fragment. Cela se produit dans **un cas sur cinq** (19% des amorces complétées et modifiées), ce qui n'est pas négligeable. Un lien a été établi entre le devenir du fragment de mot et des modifications lors de la reprise de l'énoncé (Fig.8). Les modifications antérieures sont en effet plus fréquentes dans le cas d'amorces complétées que dans celui des amorces modifiées (1/5 contre 1/10)⁵. Que des éléments de l'énoncé soient tronqués pour corriger une erreur dans l'énoncé déjà prononcé était attendu ; ce qui l'était moins c'est que cela se produise également avec des fragments de mot qui seront modifiés ; dans ce cas, le locuteur produit deux modifications lors de la reprise de son énoncé (sur le mot interrompu et sur un élément antérieur).

⁵ Khi2=5,85 ; p<.02 ; d.d.l.=1

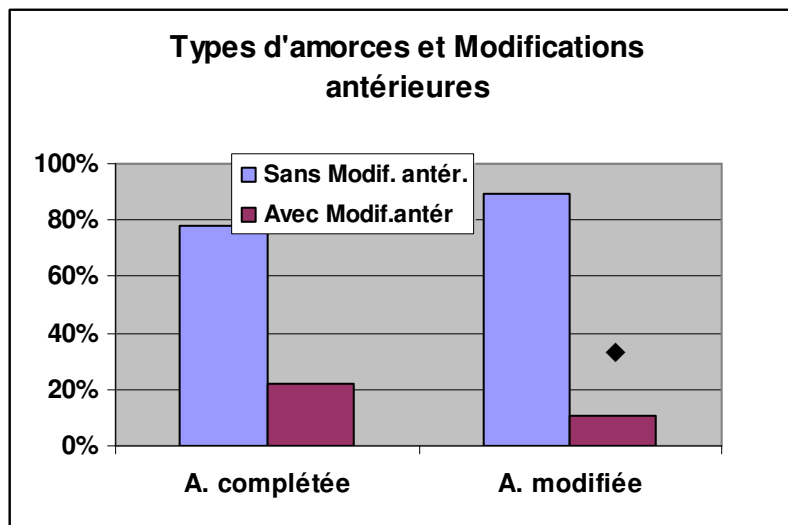


Figure 8 Présence de modifications antérieures à l'occasion d'amorces complétées et modifiées.

4.2. Les conditions de la reprise

Le résultat massif sur le devenir des amorces est que lorsque le fragment de mot est complété, il l'est toujours (et cela sans aucune exception dans nos corpus) à travers au minimum une reprise du fragment. Il n'y a jamais de simple complétude de fragment par le fragment manquant (du type : un **li-vre**). C'est un autre trait « qualitatif » qui permet de distinguer un énoncé disfluent d'un énoncé produit par une personne atteinte de « béguyage » (Zellner 1992).

Par définition, la reprise après interruption de mot n'a lieu que dans les cas d'amorces complétées ou modifiées qui sont les seuls fragments de mots suivis d'un piétinement sur la même place syntaxique. Les amorces inachevées, elles, sont justement identifiées parce que ce qui suit le fragment n'appartient pas à la même place syntaxique que lui. Le réajustement, dans ce cas-là, n'est pas une reprise mais une poursuite de l'énoncé.

Les proportions de ces trois catégories d'amorces vérifient les conclusions de Levelt (1989) sur les phénomènes d'interruption en général dans les énoncés : ces phénomènes s'accompagnent le plus souvent de reprises plutôt que d'interruption définitive de la proposition. Dans nos corpus, en effet, les locuteurs reprennent (amorces complétées et modifiées ; 78%) plus qu'ils ne poursuivent leur énoncé (amorces laissées inachevées ; 22%) après une interruption. Restait à étudier comment se fait cette reprise de l'énoncé après une interruption de mot.

Nous nous sommes intéressés à la position de l'élément interrompu et à la localisation de la reprise qui suit puisqu'on ne peut décrire la localisation des reprises après interruption de mot sans préciser auparavant la **localisation de cette interruption** qui peut avoir lieu en tête ou non du groupe nominal ou verbal⁶ :

- En tête :

ex. 11 C24aNanc 1, 3 euh la mygale **s'- s'arrime** avec ses ses crochets sur sa sa proie

ex. 12 C5bBelfo 1 parce qu'il y a **u- une** frange de jeunes qui arrivent maintenant sur euh le milieu professionnel

- Non en tête :

ex. 13 C7dBord 5,5 et et après j'ai **vou- voulu** changer

⁶ Nous nous référons ici aux notions hiérarchisées employées par Blanche-Benveniste (1997)

ex. 14 C24aNanc3 22, 3 donc les les **ani-** les **animaux** et les coraux qu'on trouve là est-ce qu'ils s-on les trouve ensemble dans la nature

Clark et Wasow (1998) ont signalé que la reprise après interruption de constituant (groupe ou syntagme) se produit le plus souvent en tête de constituant. C'est en effet le cas pour l'exemple 14 où le locuteur reprend son énoncé en « remontant » jusqu'au déterminant alors que les locuteurs 13 et 15 ne le font pas :

ex. 15 C5bBelfo, 2, 1 et il **fais-** **tenait** un une section dans un **cen-** **centre** culturel dans un petite ville du territoire ici.

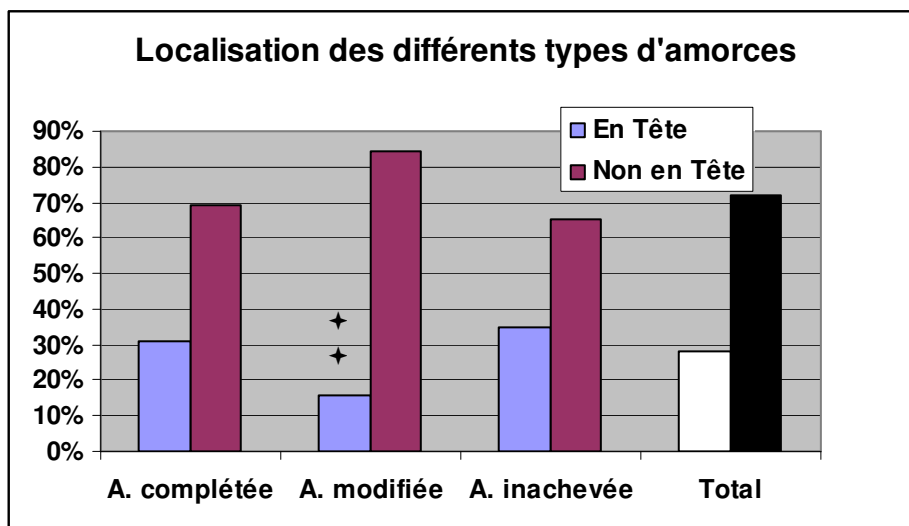
Si l'amorce de mot est déjà en tête de groupe (Syntagme verbal : *la mygale s'- s'arrime*), la localisation en tête de la reprise n'a pas la même signification que si l'amorce ne s'y trouve pas déjà (Syntagme verbal : *vous li- vous lisez*). En effet, si le fragment de mot n'est pas en tête de groupe mais que la reprise remonte jusqu'à cette tête, cela traduit une contrainte linguistique qu'il n'est pas possible de mettre en évidence lorsque l'amorce est déjà en début de groupe ; c'est pourquoi il nous a paru utile de distinguer ces deux sortes de localisation de l'amorce de mot.

4.2.1 Localisation du fragment de mot

Sur les 436⁷ interruptions de mots où il a été possible de déterminer si elles avaient lieu en tête de groupe ou non (Fig.9), plus des deux tiers (72%) ne se produisent pas en tête de groupe (nominal ou verbal ; pas de différence). Ce résultat n'est pas surprenant puisque la grande majorité de ces disfluences touchent, on l'a vu le lexique, item qui en français est rarement en tête de groupe mais précédé d'un déterminant ou d'un pronom.

Nous avons constaté (Fig.9) que la localisation de la perturbation ne permet pas de prédire si il y aura reprise (Amorce complétée) ou simple poursuite de l'énoncé (Amorce inachevée) : en effet, quelle que soit la localisation de l'amorce (en tête ou non), on trouve une proportion semblable (35%) de fragments inachevés (sans reprise, donc) et de fragments qui seront complétés (31%)⁸.

Seule la proportion de fragments qui seront modifiés diffère significativement⁹. Ces amorces sont plus nombreuses lorsque la perturbation n'est pas située en tête de groupe. Ce résultat tient sans doute au fait que c'est l'amorce lexicale et non le déterminant ou le pronom qui est plus souvent remplacée (le nom ou le verbe se trouvent rarement en tête de syntagme).



⁷ Dans 5 cas il n'est pas possible de déterminer si le fragment de mot est en tête de syntagme ou non.

⁸ La différence n'est pas significative.

⁹ ($\chi^2 = 10,46$; $p < .01$; d.d.l.=2)

Figure 9 Localisation (en tête ou non de syntagme) des différents types d'amorces de mots

4.2.2. Emplacement de la reprise suivant un fragment de mot qui n'est pas en tête de groupe.

Les reprises observées à la suite des amorces (qui seront complétées ou modifiées) peuvent être regroupées dans trois catégories :

*Les reprises « **minimales** » qui ne remontent pas jusqu'au début du groupe :

ex. 16 CorpusEDF, 12,6 la note est est salée hein la note **énergé- énergétique**

*Les reprises « **Début du Groupe** » qui remontent jusqu'au début du Groupe nominal ou verbal :

ex. 17 C5cBelfo, 2,2 alors ce monsieur s'est approché et il a dit **vous li- vous lisez** l'Est Républicain donc vous êtes de l'Est

*Les reprises « **Plus** » qui remontent au delà du groupe :

ex. 18 Corpus EDF, 8,38 mais *l'eau est de vingt mè-* ***l'entrant est de vingt mètres*** cubes

Ce type de disflueance confirme les observations faites par Blanche-Benveniste (1997 p 48) et les résultats signalés sur tous les types de disflueance par Clark et Wasow (1998) : après une interruption, le locuteur reprend généralement depuis le début d'un syntagme. Dans notre étude (Fig.10), les reprises après troncation de mot ont lieu le plus souvent au début du groupe (63%). Il reste qu'une part non négligeable des reprises ne se conforme pas à ce modèle puisque **37%** d'entre elles ne se font pas au début du groupe nominal ou verbal. Certaines, les reprises « minimales » (24%) ne remontent pas au début du syntagme et d'autres, les reprises « Plus » (13%), au contraire, dépassent cette limite.

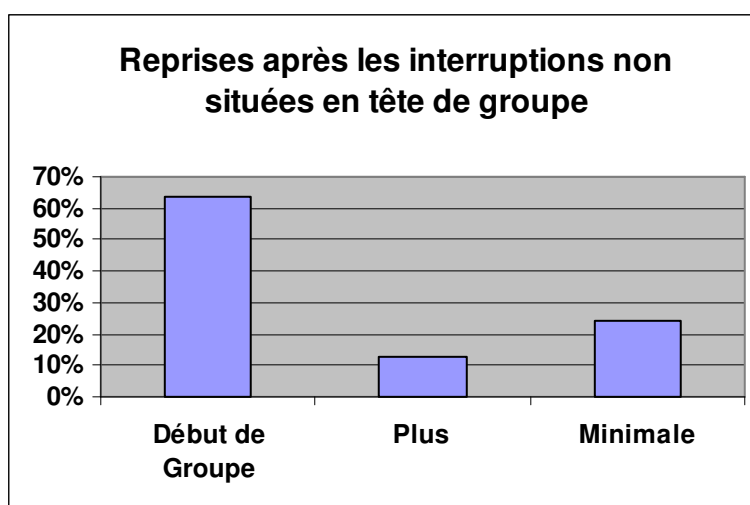


Figure 10 Types de reprises après des interruptions non situées en tête de syntagme

4.2.3. Emplacement de la reprise qui suit un fragment de mot situé en tête de groupe.

Dans nos corpus, seuls 28% des amorces de mots sont en tête de groupe (Fig. 9). Comment s'effectue la reprise dans ces cas-là ? Va-t-on trouver des reprises « Plus » ? Les amorces complétées et modifiées (qui sont donc suivies d'une reprise) sont, dans leur grande majorité (92%), suivies d'un réajustement localisé sur la place du segment de mot, donc en tête de groupe.

ex. 19 C5cBelfo 16,1 j'ai d'abord été assez suffoquée j'ai eu **div- diverses** réactions

ex. 20 C5cBelfo, 10, 4 alors euh comme les enfants **deman- demandaient** des des renseignements sur les mots

Rien ne s'oppose cependant à ce que la reprise ait lieu plus avant dans l'énoncé (reprise « Plus »). Quelques unes (6%), en effet, dépassent cette limite ([2] et remontent sur des éléments antérieurs ([1] :

ex. 21 C6cBesan, 6, 4 euh dans la mise en pratique [1]je vois pas [2c- [1]je verrais pas [2**comment** faire correspondre les choses

ex. 22 Tropr102, 3,1, [1]je crois que [2c- + [1]je crois que [2c'**est** la famille des N. mais j'en suis pas sûr

On constate (Fig.11) que le locuteur sort, aussi souvent, des limites du syntagme lorsqu'il reprend plus avant son énoncé (reprise « Plus ») que l'interruption ait lieu en tête ou non du syntagme. Ce type de reprise ne se produit que dans 10% des cas environ et ne dépend pas de la localisation de l'amorce de mot.

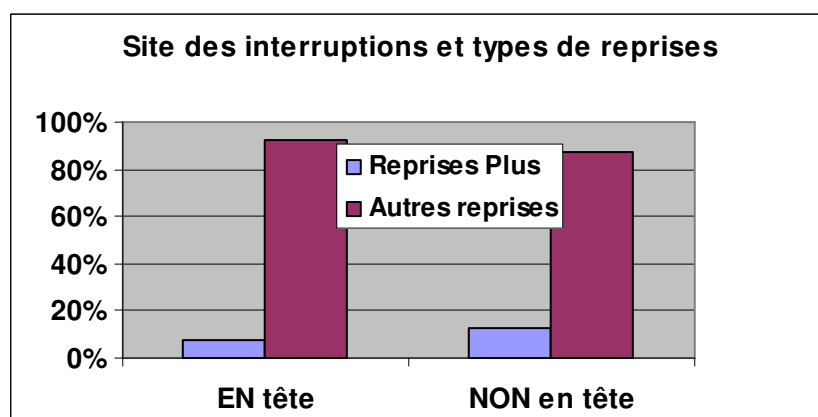


Figure 11 Localisation de l'amorce et types de reprises.

4.2.4. Emplacement de la reprise et types d'amorces

Le type d'amorce est un paramètre qui exerce une influence sur la localisation de la reprise qui suit. Le locuteur qui reprend son énoncé après une interruption de mot ne le fait pas de la même façon si le fragment est complété, modifié ou laissé inachevé (Fig.12). Pour compléter son amorce, le locuteur reprend significativement plus souvent¹⁰ son énoncé au début du syntagme que lorsqu'il la modifie. Dans ce dernier cas, il augmente la fréquence de deux stratégies : reprise minimale et reprise d'éléments antérieurs au syntagme concerné par l'interruption (Reprise « Plus »)

¹⁰ Khi2=11,64 ; p<.001 ; d.d.l.=2

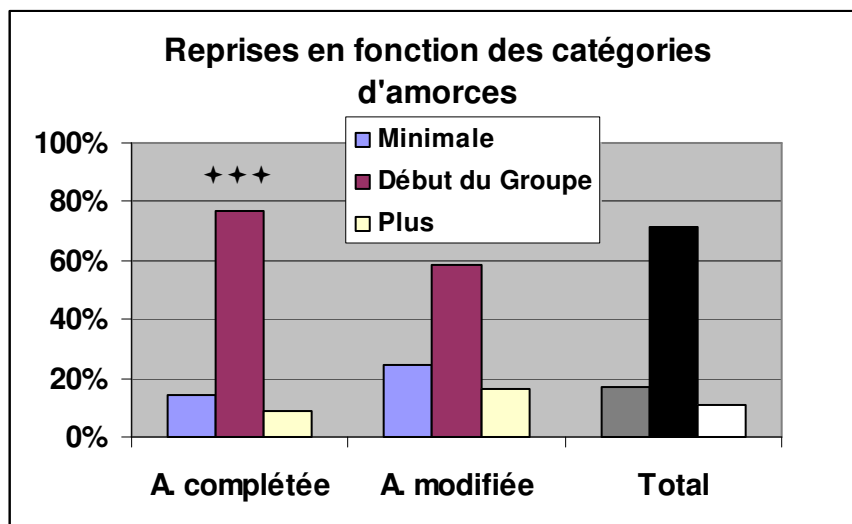


Figure 12 Répartition des types de reprises dans les différentes catégories d'amorces

4.3. Troncations et rupture syntaxique

Nous avons vu que la troncation peut s'accompagner d'un piétinement syntaxique (amorces complétées et modifiées) ou non (amorces inachevées). Le plus souvent, même lorsqu'il s'agit d'un segment qui sera laissé inachevé, cette interruption, ne s'oppose pas à la poursuite de l'énoncé c'est à dire à une continuité syntaxique. Le locuteur continue de développer le syntagme suspendu et produit d'autres arguments pour le même verbe recteur :

ex. 23 C5cBelfo, 2,2 ce monsieur s'est approché et il a dit vous **li-** vous **lisez** l'Est Républicain donc vous êtes de l'Est

ex. 24 C5aBelfo, 19, 2 euh leur faire prendre conscience qu'ils sont tout à fait capables **de f- de poser** les choses eux-mêmes

ex. 25 Tropr102, 5,3 le parrain **ita-** enfin de la mafia italienne ça ne peut être qu'un sicilien et toujours un sicilien

Dans d'autres cas au contraire, ce qui suit le fragment ou son réajustement correspond à **une rupture dans la construction syntaxique** (c'est à dire que ce qui suit n'appartient plus à la rection du verbe) :

ex. 26 Pariscentre, 1, 1 euh **c'ét- c'était** ça m'a demandé beaucoup d'effort de compréhension

ex. 27 C41cStg 20, 5 donc c'est mh le comptage du + **le ni-** comment dire ça **le léchage** euh le mh + quand le l'animal vient euh euh pff boire donc de compter le nombre de fois où il vient boire

ex. 28 C7bBord, 2,1 pff tout est allé très vite en fait pour moi c'est j'ai décidé j'ai **lai-** c'est-à-dire dans ma tête c'était un vaste projet

On constate que les troncations de mots (toutes confondues) sont, dans la grande majorité des cas (88%) suivies d'une poursuite de la proposition. Cette interruption ne provoque que dans 12% des cas seulement une rupture de construction verbale. Cependant, si des exemples de ces cas sont observés dans les trois types de fragments de mots, ils le sont dans des proportions considérablement différentes (Fig. 13)¹¹. A quelques exceptions près, la majorité des amorces de mots (qui seront complétées ou modifiées : 78% des amorces de mots donc) ne sont pas suivies

¹¹ Khi2=138,78 ; p<.0001 ; d.d.l.=2

d'une rupture syntaxique. Il n'est pas de même pour l'amorce qui est laissée inachevée (22%) : près de la moitié (46%) de ces troncations sont suivies d'une rupture de construction verbale.

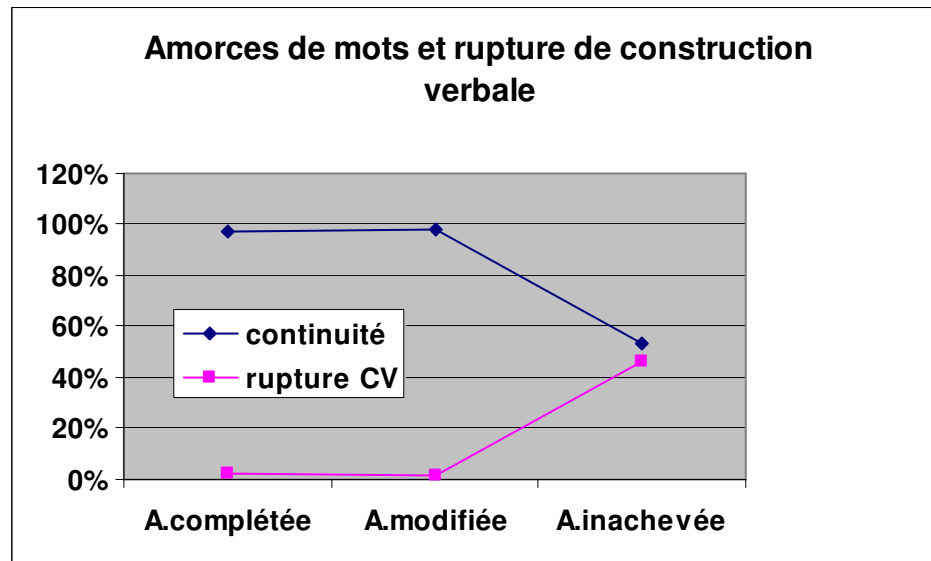


Figure 13. Rupture de construction verbale après des amorces de mots complétées, modifiées et inachevées

Ainsi, lors de la reprise, les éléments trouvés après le fragment complété ou modifié enrichissent, dans la très grande majorité des cas, le contexte de la réaction verbale. Il n'en est pas de même pour les fragments de mots qui seront laissés inachevés. Dans 46% des cas, ce qui suit le fragment inachevé montre qu'il n'y a pas d'enrichissement du contexte verbal car la troncation du mot interrompt, non seulement le syntagme, mais aussi la proposition.

Cette troncation suivie de rupture syntaxique est-elle sous la dépendance de paramètres tels que la localisation de l'amorce et le type de mots touchés par l'interruption ?

La localisation de l'amorce de mot par rapport au verbe recteur a une influence sur l'impossibilité pour le locuteur de maintenir la continuité de son énoncé (Fig. 14). Les cas de rupture de construction verbale sont deux fois plus nombreux avant et sur le verbe qu'après le verbe. Autrement dit, plus la troncation a lieu tôt (avant et sur le verbe), plus l'énoncé risque de s'interrompre¹². En revanche, les cas de continuité syntaxique sont aussi nombreux, quelle que soit la localisation du fragment de mot.

¹² $\chi^2=6,44$; $p<.02$; d.d.l.=1

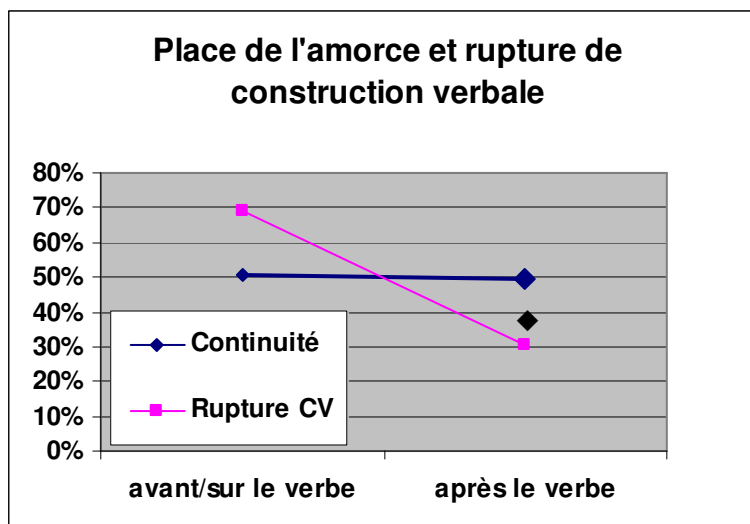


Figure 14 Influence de la localisation de l'amorce sur la poursuite de l'énoncé (continuité syntaxique ou rupture de construction verbale)

Comme ces ruptures sont réservées aux cas d'interruptions laissées inachevées, leur plus grande fréquence sur la zone avant/sur le verbe pourrait être dû au fait que les amorces laissées inachevées se produisent plus souvent dans cette zone. Nous vérifions qu'il n'en est rien (Fig. 15)¹³. Lorsque l'énoncé se poursuit sans rupture syntaxique, la répartition des amorces inachevées par rapport au verbe est la suivante : 60% après le verbe et 40% avant ou sur le verbe.

Lorsqu'il y a rupture syntaxique, ces proportions sont inverses : 24% des amorces inachevées sont après le verbe et 76% le sont avant ou sur le verbe. Il semble donc que, chez les amorces de mots laissées inachevées, la localisation de l'interruption avant ou sur le verbe permette d'anticiper un risque plus grand de rupture de construction verbale.

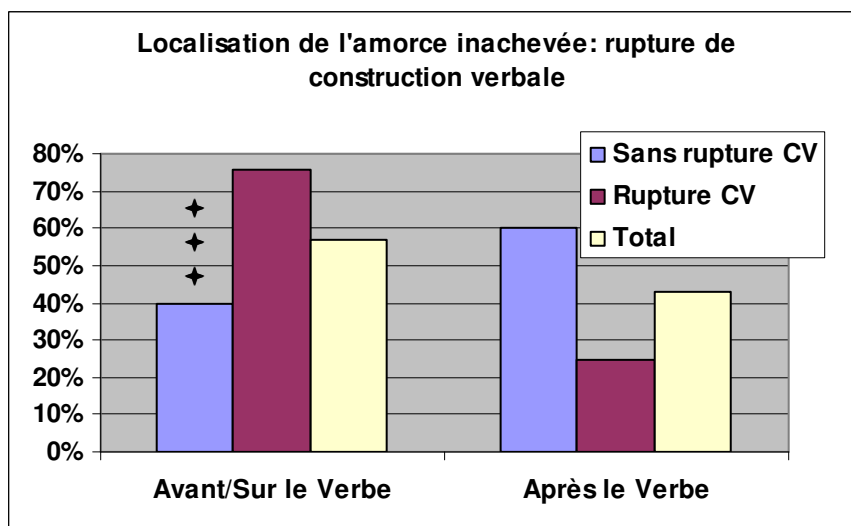


Figure 15 Comparaison des ruptures syntaxiques produites à la suite des amorces de mots laissées inachevées en fonction de leur localisation dans la proposition.

Par ailleurs, nous avons vu (Fig. 3) que les troncations laissées inachevées touchent significativement plus souvent les mots-outils (36%) si on les compare aux troncations complétées

¹³ $\chi^2=12,20$; $p<.001$; d.d.l.=1

(23%) ou modifiées (15%). Cette spécificité des amorces laissées inachevées pourrait être également liée à une propension de ces troncations à être suivies d'une rupture syntaxique.

Le type de mots semble, en effet, avoir une influence sur la possibilité qu'a le locuteur de pouvoir compléter sa proposition si l'amorce est laissée inachevée (Fig.16). Les mots-outils interrompus et laissés inachevés sont significativement¹⁴ plus souvent que les mots-pleins suivis d'une rupture de la construction verbale (65% contre 33%).

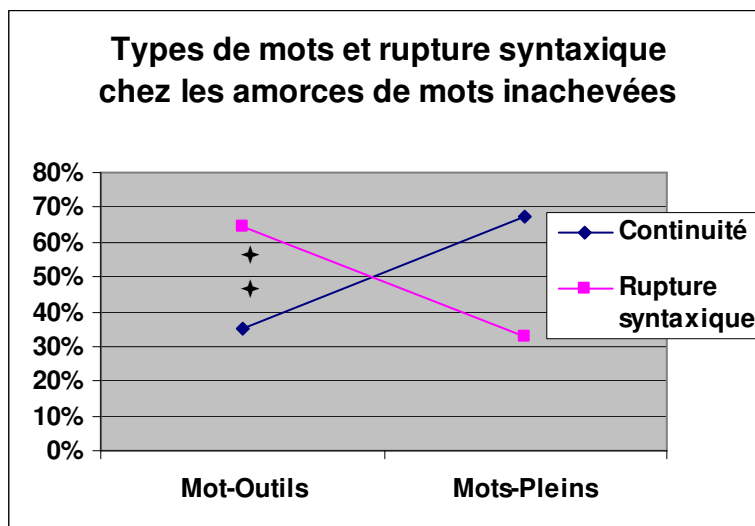


Figure 16 Types de mots et ruptures de construction verbale chez les amorces de mots inachevées.

5. Conclusion

Les productions d'amorces de mots décrites dans les corpus de français parlé « standard », montrent que ces phénomènes sont très fréquents et participent tous activement à l'élaboration de l'énoncé à travers des piétinements successifs sur les places syntaxiques. Au moins la moitié des troncations involontaires (celles qui seront complétées) dans les énoncés peuvent être qualifiées d'hésitations. Si le statut des fragments de mots qui seront suivis d'une modification paraît être clairement une erreur, celui des fragments de mots laissés inachevés est moins clair puisque le locuteur ne confirme ni n'infirme sa production par une reprise (qui seule, permet de savoir si il s'agit d'une erreur ou d'une hésitation).

Répétitions et interruptions de mots sont deux phénomènes distincts qui n'affectent pas les mêmes places syntaxiques. On répète plus des mots-outils que des mots lexicaux et on s'interrompt plutôt au milieu d'un item lexical que d'un mot-outil. Chez les bégues, Zellner (1992, p482) constate, au contraire, que lors d'une hésitation (et non d'un bégaiement), « quelque soit le taux de disfluences, les mots fonction monosyllabiques – qui permettent de structurer le discours – ont tendance à subir plus d'accidentes relativement aux autres mots ». D'autre part, nous n'avons jamais observé dans nos corpus (comme on peut l'observer dans 40% des disfluences produites par des bégues ; Zellner, 1992 ; Bensalah, 1997, van Hout 2002) de « syllabes progressivement produites en plusieurs émissions » ni de complétude du mot amorcé sans qu'il y ait reprise du début du mot. Parfois, très rarement (une fois toutes les 33 min), une amorce de mot est répétée avant que le mot ne soit finalement complétée (dans 80% des cas).

Nous avons étudié les réajustements auxquels le locuteur procède après s'être interrompu au milieu d'un mot d'un double point de vue : d'une part, les types d'insertions (30% des cas) et d'autre part les reprises de l'énoncé (75% des cas).

¹⁴ khi2=8,85 ; p<.01 ; d.d.l.=1

Deux sortes d'insertions ont été distinguées : les insertions non liées syntaxiquement à l'énoncé et celles qui le sont. Les premières sont observées surtout immédiatement après l'interruption parfois (rarement) lorsque la reprise a commencé. Elles sont deux fois plus nombreuses lorsque le fragment de mot est laissé inachevé. On peut émettre l'hypothèse qu'elles viennent prendre la place d'une reprise de l'énoncé déjà prononcé ce qui va permettre au locuteur de poursuivre son énoncé sans avoir complété son syntagme. Les insertions liées syntaxiquement sont réservées au moment où la reprise commence. Ces dernières modulations, réservées aux fragments complétés ou modifiés, consistent essentiellement en des ajouts ou des remplacements de mots, les éléments retirés étant peu fréquents. La présence de ces « modifications antérieures » lors de la reprise justifie le doute sur la localisation de l'hésitation. Dans un cas sur cinq, le moment de l'interruption n'est pas celui de l'erreur mais celui de sa perception par le locuteur qui s'interrompt alors y compris au milieu d'un mot. Nous avons constaté, de plus, que même lorsque le fragment de mot est modifié (donc, apparemment une erreur), il est, dans un cas sur dix, le signe indicateur d'une erreur prononcée auparavant et donc d'une modification antérieure.

Les interruptions de mots se produisant dans la grande majorité des cas sur du lexique, elles ne sont donc pas en tête de groupe nominal ou verbal. La reprise peut se faire sur le mot même interrompu, remonter jusqu'au début du groupe (nominal ou verbal) ou même dépasser cette limite. Si les cas les plus nombreux sont bien (comme l'avaient signalé Glark et Wasow, 1998) ceux où la reprise de l'énoncé remonte jusqu'au début du groupe, un tiers des reprises échappent à ce schéma, ce qui n'est pas négligeable. Près de 24% des reprises sont en effet « minimales » et ne remontent pas jusqu'au début du groupe. Les autres au contraire dépassent la limite du Groupe nominal ou verbal. On constate que les reprises (6 à 10 %) qui dépassent les limites du groupe où a lieu l'interruption ne dépendent pas de la localisation de cette dernière. En revanche, lorsque le locuteur complète le fragment de mots interrompu, il remonte plus volontiers au début du syntagme qu'il ne le fait lorsqu'il modifie le fragment interrompu. Dans le cas d'amorce modifiée, il adopte plus souvent les stratégies « minimale » ou « Plus » indifféremment.

Contrairement à ce qui se passe pour les amorces complétées et modifiées qui ne provoquent qu'exceptionnellement un arrêt du déroulement de la proposition, dans **46%** des cas, ce qui suit le **fragment inachevé** interrompt la proposition. Ce sont sans doute ces cas de rupture de la construction verbale qui sur le plan perceptif provoque cette impression de ratés dans l'énoncé. De nos analyses sur les interruptions de mots laissés inachevés et suivies d'une rupture de construction verbale, il semble possible de conclure que deux paramètres semblent influencer le devenir de l'énoncé : il y a plus de chance que le locuteur ne poursuive pas sa proposition, d'une part, si il s'interrompt sur un élément avant ou sur le verbe et, d'autre part, si cet élément est un mot-outil.

6. References

- Blanche-Benveniste C., 1997, *Approches de la langue parlée en français*. Edition Ophrys, Paris.
- Candéa M., 2000, Contribution à l'étude des pauses silencieuses et des phénomènes dits « d'hésitation » en français oral spontané. Étude sur un corpus de récits en classe de français. Thèse d'État, Université Paris III (Sorbonne Nouvelle).
- Cappeau P., 1998, Quelques mots sur quelques bribes liées au genre. In: Bilger M, Van den Eynde & Gadet F, (Eds) *Analyse linguistique et approches de l'oral. Recueil d'études offert en hommage à Claire Blanche-Benveniste*. Peeters, Leuven, Paris, 301-311.
- Clark et Wasow, 1998, Repeating words in spontaneous speech, *Cognitive Psychology*, 37, p 201-242
- Duez D., 1997, La signification des pauses dans la production et perception de la parole., *Revue Parole*, 3-4, 275-299.
- Duez D., 2001, Signification des hésitations dans la production et la perception de la parole spontanée., *Revue Parole*, 17-18-19, p 113-138.
- Henry S., 2002a, Étude des répétitions en français parlé spontané pour les technologies de la parole, *RECITAL'02*. Nancy (France) : ATALA, 467-476.
- Henry S. & Pallaud B., 2003, Word fragments and repeats in spontaneous spoken French. In: Eklund R., *Disfluency in Spontaneous Speech Workshop*, Proceedings of DiSS'03, 5-8 September 2003, Göteborg University, Sweden, p 77-80.
- Henry S. & Pallaud B., 2004 c, Amorce de mots et répétitions dans les énoncés oraux, *Recherche Sur le Français Parlé*, 18, 201-229.
- Van Hout A. 2002, *Les bégaiements. Histoire, psychologie, évaluation, variétés, traitements*, Paris, Masson, 2^{ème} édition. (1997)
- Jeanjean C., 1984, " Les ratés c'est fa- fabuleux". Etude syntaxique des amorces et des répétitions., *LINX*, 10, *Syntaxe et discours*, p. 171-177.
- Levelt W.J.M., 1983, Monitoring and self-repair in speech. *Cognition*, 14, 41-104.
- Levelt W.J.M., 1989, *Speaking. From intention to articulation*. MIT Press, Cambridge.
- Pallaud B., 2002 a, Les amorces de mots comme faits autonymiques en langage oral, *Recherches Sur le Français Parlé*, 17, 79-102.
- Pallaud B., 2003 a, Achoppements dans les énoncés de français oral et sujets syntaxiques. In Merle J.M. (Ed.), *Le Sujet* ; Paris : Éditions Ophrys, *Faits de Langue*, p 91-104.
- Pallaud B. & Henry S. , 2004, Amorce de mots et répétitions : des hésitations plus que des erreurs en français parlé. In *Le poids des mots. Actes des 7èmes Journées Internationales d'Analyse statistique des Données Textuelles*, Louvain-la-Neuve, 10-12 mars 2004. Louvain, PUL, vol 2, 848-858.
- Pallaud B. & Henry S., 2005, Les effets langagiers des amorces de mots et des répétitions dans les énoncés de français parlé spontané., *Perturbations et réajustements. Langue et Langage*, (Réseau Grand-Est des Sciences Cognitives (COGNIEST), Haguenau, 1-3 Décembre 2004.(à paraître)
- Padeloup V., 1992, A Prosodic Model for French Text-to-Speech Synthesis : A Psycholinguistic Approach. In BAILLY, Gérard; BENOIT, Christian; SAWALLIS, Thomas R. (eds), *Talking Machines. Theories, Models, And Designs*, pp. 335-348
- Schegloff E., Jefferson H. and Sachs H., 1977, The preference for self-correction in the organization of repair in conversation., *Language*, 53, 2, p 351-382.
- Shriberg E. & Stolcke A., 1998. [How Far Do Speakers Back Up In Repairs? A Quantitative Model](#). *Proc. Intl. Conf. on Spoken Language Processing*, vol. 5, pp. 2183-2186, Sydney, Australia.
- Shriberg E. (1999). Phonetic Consequences of Speech Disfluency. Symposium on The Phonetics of Spontaneous Speech (S. Greenberg and P. Keating, organizers), *Proc. International Congress of Phonetic Sciences*, vol. 1, pp. 619-622, San Francisco.

Zellner, B. (1992). Le bé- bégayage et euh ... l'hésitation en français spontané. Actes des 19 èmes Journées d'Études sur la Parole, J.E.P. Bruxelles, pp. 481-487.